

**DIFFERENTIAL MEASUREMENT ERROR ACROSS TIME AND TREATMENT
GROUPS IN SELF-REPORTED NUTRITION DATA**

by

Adam R. Pittman

A thesis submitted to Johns Hopkins University in conformity with the requirements for the
degree of Master of Science

Baltimore, Maryland

August 2019

ABSTRACT

In dietary observational studies and randomized controlled trials, researchers often try to ascertain the relationship between nutrient intake and biological outcomes, or the effect of interventions on intake and other outcomes. From a study design perspective, this requires capturing true participant intake, but properly doing so is (nearly) impossible. Intake measurements often rely heavily on self-reported nutrition measurements which are an easy and cost-effective proxy to implement in studies, but may not accurately reflect true nutrient intake. Less frequently, intake measurement relies on measured biological components (such as blood or urine) known as biomarkers. Biomarkers, considered to be the “gold standard”, are more resource and financially intensive, but better represent true intake.

When studies contain both self-reported and biomarker nutrient values (which happens sparingly), researchers can model the measurement error structure for self-reporting errors and attempt to produce less biased results for calculating true but unobservable nutrient intake. Previous measurement error work that investigates the relationship between biomarker and self-reported levels has typically been at a single time point, in a single treatment group, or with respect to basic patient demographics. Few studies have examined the measurement error structure in longitudinal studies, where nutrient intake and self-reported values may change over the course of a study, and by treatment exposure.

Using two longitudinal randomized controlled trials with internal validation data (urine biomarkers and self-reported values), we examine how self-reported sodium error changes as a function of time and/or treatment assignment by comparing it to measured urine sodium.

We find that although true sodium consumption changes across time and treatment group, there is essentially no evidence that the measurement error varies across time or treatment groups. While researchers should consider the effects of time and treatment status when designing longitudinal studies, more evidence is needed on how measurement error changes with regards to time and treatment.

Thesis Advisor: Elizabeth A. Stuart

Thesis Reader: John McGready

ACKNOWLEDGEMENTS

I would like to thank all my graduate school friends and professors, who made my time here an amazing, enriching experience. A special thanks to,

My advisor, Liz Stuart, who always supported me and provided invaluable feedback on all my work.

John McGready, whose door was always open to me.

Juned Siddique, who helped make this thesis possible.

And Karen Bandeen-Roche, the most wonderful person I have ever met.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
List of Tables	v
List of Figures	vi
1. Introduction	1
1.1 Nutrition Data	1
1.2 Measurement Error Models	3
1.3 Correcting for Measurement Error	5
1.4 Application	7
2. Data	9
2.1 Trials of Hypertension Prevention	9
2.2 PREMIER	10
3. Methods	13
3.1 Covariate Modelling	15
4. Results.	17
4.1 Descriptive Results	17
4.2 Regression Results	20
4.3 Covariate Results	22
5. Discussion	24
5.1 Limitations	25
6. Supplementary	27
6.1 Covariate Models	27
6.2 Covariate Output	29
7. References	33
Curriculum Vitae	35

LIST OF TABLES

Table 1. Study Characteristics	12
Table 2. Mean log values by time and treatment in the TOHP dataset.	17
Table 3. Mean log values by time and treatment in PREMIER.	20
Table 4. Standardized Regression output from model (3)	21
Table 5. Percent Missing by Time/Treatment in each study	26
Supplementary Table 1. Standardized Regression output from BMI model (3)	30
Supplementary Table 2. Standardized Regression output from sex model (3)	31

LIST OF FIGURES

Figure 1. Measured Urine vs. Self-Reported intake in each study by time and treatment . **17**

Supplementary Figure 1. Measured urine vs. self-report by sex in both datasets **29**

1. INTRODUCTION

1.1 Nutrition Data

As chronic disease remains a steadfast issue in public health, it is natural to focus on lifestyle interventions to identify promising programs to improve overall population health. In dietetics, researchers apply such interventions to understand how improving nutrition habits will affect individual well-being. In a randomized control trial (RCT), scientists use chance to decide who receives an intervention(s) and who is placed in a control group(s) to compare dietary approaches such as limiting consumption of hypertensive causing nutrients, or meeting with a registered dietician (treatment) to an individual with no counseling (control) and assess how these interventions or dietary changes influence health. From this information, we discern what dietary factors lead to increased (or decreased) risk for a disease outcome.

Obtaining accurate measures of nutrient intake is important for understanding the diet and chronic disease relationship, or when studying participant's eating behaviors, yet properly measuring food consumption with high accuracy can be difficult. Direct nutrient intake is rarely observed, and in dietetic studies, researchers frequently resort to two methods to measure nutrient intake; biomarkers and self-reported methods.

Biomarkers are biologic components from participants, such as blood, urine, or hair which contain information about a person's nutrient levels. Biomarkers are useful because they objectively measure intake and are considered the "ideal". Therefore, biomarkers may be closer to the "truth" than self-reported methods, and hence estimate a person's nutrient intake.

Unfortunately, biomarkers are often expensive, invasive, and/or difficult to implement into a study (Kirkpatrick 2018). They place potentially greater burden on trial participants than

self-report measures, which may discourage people from taking part of a lifestyle trial. Thus, there are concerns that biomarkers can contribute to poor trial adherence and missing data problems (i.e., that participants will drop out of the study because of the hassle or invasiveness of the biomarker collection) (Prentice 2002). For these reasons, it is often infeasible to capture biomarker data over time in many studies.

On the other hand, self-reported data comes directly from the participant, where they inform study officials about their food consumption. This often takes the form of food frequency questionnaires (FFQ), where participants fill out a survey about their eating habits or 24-hour diet recall, where people report everything consumed over the previous day. Self-reported methods are more frequently implemented than biomarker measurements since these are likely easier, cheaper, and more convenient for the participant (Kirkpatrick 2017).

These two methods act as “proxy” measurements of true intake, because they can be representative, but are potentially imprecise versions of the truth and experience two main types of error: systematic and random. Systematic error, or bias, consistently departs from the truth in the same direction (i.e., always higher or lower), and can be hard to detect and analyzed statistically (NIH 2014). Systematic errors can decrease the accuracy of measurements and create potentially erroneous conclusions about the relationship between food intake or nutrients and nutrition-related diseases (Gibson 2017). Random error can create variability in the measurements, which may reduce precision, resulting in a loss of statistical power. However, random errors can be more easily corrected with statistical methodology (Thompson 2016). This gives rise to measurement error, the difference between “true” nutritional exposure and “observed” intake.

Biomarkers are considered the “gold standard” compared to self-report (and other nutrient reporting methods), because while biomarkers have a random error component, they potentially experience less systematic error than self-reported information. Self-reported measures are more susceptible to both random and systematic measurement error because, even with the best intentions, people improperly remember the foods they ate (recall bias) or lie about true consumption to appear healthier to researchers (desirability bias) (Espeland 2001).

Given these measurement challenges in nutrition and many other fields, a body of research around methods to deal with measurement error has grown up in statistics. We will now formally define measurement error, and some developed methods to correct for it.

1.2 Measurement Error Models

Biomarkers and self-reported measurements, are some examples of measurement error sources in nutritional epidemiology. However, many other similar instruments exist in different disciplines such as air pollution monitors (environmental health) or chemical balances (chemistry). For this paper, we focus on dietetics and nutrient applications, but these concepts can be extrapolated to other fields which experience measurement error.

We start with *classical measurement error* (Carroll 2006). Let \mathbf{X}_i , (a vector of length N , where N = number of trial participants), be the true, but unobservable value of intake for person i . As researchers, we can only observe \mathbf{W}_i (where $\mathbf{W}_i = (w_1, \dots, w_n)$), a proxy measurement(s) of \mathbf{X}_i , such as self-reported data, in person i . Since \mathbf{W}_i doesn't perfectly match the true value \mathbf{X}_i , some error is created such that:

$$\mathbf{W}_i = \mathbf{X}_i + \boldsymbol{\varepsilon}_i.$$

Where $\epsilon_i | \mathbf{X}_i \sim \text{Normal}(0, \sigma_E^2)$. This means the observed dose, \mathbf{W}_i , has higher variability than the true dose, \mathbf{X}_i (Carroll 2006). In dietary studies, self-reported values generally have greater variability than biomarker values, and classical measurement error is a justifiable model in this situation.

We can expand the classical measurement error model to include proxy measurements \mathbf{W} , and additional covariates \mathbf{Z} , assumed to be recorded without any error, like age or gender. Let \mathbf{Z} be a $N \times K$ matrix, where N is the number of trial participants, and K is the number of recorded covariates for each participant. Now our equation resembles:

$$\mathbf{W} = \mathbf{X} + \mathbf{Z} + \mathbf{XZ} + \epsilon, E(\epsilon | \mathbf{X}, \mathbf{Z}) = 0.$$

This equation says, proxy measurement \mathbf{W} , is affected by true intake \mathbf{X} , and covariates \mathbf{Z} , i.e., BMI, gender, or age might all affect self-reported amounts. While presented as an expansion of classical models, many measurement error models can be expanded to include covariate terms.

Another example of measurement error modelling is *Berkson measurement error* (Berkson 1950), which is more common in other disciplines, such as occupational epidemiology. Keeping the same notation as above, \mathbf{W}_i now represents a known quantity, but true intake, \mathbf{X}_i , is still unknown such that:

$$\mathbf{X}_i = \mathbf{W}_i + \epsilon_i.$$

Where $E(\epsilon_i | \mathbf{W}_i) = 0$. This model now assumes the truth has more variability than the estimated dose (Carroll 2006). An example would be when people get the same known radiation exposure level \mathbf{W}_i , but the total amount of radiation levels in the body, \mathbf{X}_i , varies across people (Carroll 2011).

These examples are by no means an exhaustive list of measurement error models. In the models above, we assumed the truth and observed values were continuous variables with a linear relationship. For nonlinear relationships, see Carroll et al. (2006.) Additionally, measurement error might occur in discrete outcomes, this is known as *misclassification*. Misclassification, commonly referred to as sensitivity and specificity in epidemiology, occurs when an outcome is incorrectly diagnosed. One example is when an unaffected person tests positive for a disease they don't have. See Braga-Neto (2009) for misclassification in genetic research or Friedman et al. (2000) for machine learning applications.

Finally, any measurement error model may be subject to non-differential or differential error. Non-differential occurs if measurement error is the same between exposed and unexposed groups or between subject with and without a specified health outcome (sick vs. healthy). Differential measurement error occurs when measurement error is different between treatment groups or different for those with an outcome of interest compared to those without the outcome of interest (Alexander 2015).

1.3 Correcting for Measurement Error

We'll now transition from describing the measurement error dilemma, to examining ways to correct for it. We center our applications around measurement error structure in dietetic studies, but many of these concepts are applicable to other disciplines and measurement error models.

Regression Calibration is one method for correcting measurement error (Carroll et al. 1995). To build this model, it is necessary to use a dataset which contains information on both participants' self-reported intake and recovery biomarker intake. By assuming biomarkers are

commensurate with the “true” intake values, \mathbf{X} , we attempt to produce less biased results estimating diet-outcome associations. As mentioned earlier, datasets with both biomarker and self-reported values are relatively rare because of implementation challenges in large longitudinal studies.

Although \mathbf{X} cannot be observed directly, researchers can use a biomarker value \mathbf{T} , into the equation as a substitute and make an assumption that $E(\mathbf{X}|\mathbf{W}, \mathbf{Z})$ closely resembles $E(\mathbf{T}|\mathbf{W}, \mathbf{Z})$. Biomarkers aren’t perfect as they may have their own measurement error, but they can be used as a more accurate measure of \mathbf{X} , compared to other observable measurements. This substitution, while not perfect, allows researchers to investigate the relationship between the biomarker (acting as the “truth”) and proxy self-report measurement.

$$\mathbf{X} = \beta_0 + \beta_1' \mathbf{W} + \beta_2' \mathbf{Z} + \beta_3' \mathbf{W} * \mathbf{Z} + \epsilon, \quad E(\epsilon | \mathbf{W}, \mathbf{Z}) = 0.$$

OR

$$\mathbf{T} = \beta_0 + \beta_1' \mathbf{W} + \beta_2' \mathbf{Z} + \beta_3' \mathbf{W} * \mathbf{Z} + \epsilon, \quad E(\epsilon | \mathbf{W}, \mathbf{Z}) = 0.$$

One example of a calibration study in dietetics is Neuhouwer (2008), using data from the Women's Health Initiative Dietary Modification Trial. In this paper, researchers used a calibration equation to show that age, BMI, race, and treatment assignment (albeit at one time point) played a role in self-reported values.

Simulation Extrapolation, or SIMEX, is another, possibly more robust, method for correcting measurement error (Cook 1994). SIMEX simulates additional datasets, each with increasing user-induced amounts of measurement error on fitted coefficients. Each new simulation creates larger measurement error, and then tries to extrapolate the trends from

increasing measurement error datasets into the original model (Carroll 2006) to see which coefficients contribute the most to variability in the data.

From Hardin (2003),

“Regression calibration attempts to estimate the unknown covariate and then run the analysis of interest using this linear approximant in place of the unknown covariate. SIMEX, on the other hand, simulates data in order to see the effect of measurement error on the fitted coefficients so that we can extrapolate back to the results we would have if the covariate were known”

See Shang (2012), for a more in-depth discussion, in which the author used SIMEX to reduce measurement error in predicting student growth percentiles.

Again, this is not an exhaustive collection of measurement error correction methods. Other examples include sensitivity analysis, which investigates the plausibility of certain assumptions when modelling measurement error (Siddique et al. 2018), or Bayesian methods, which utilize simulations and Markov Chain Monte Carlo concepts (Natarajan et al. 2010).

For the purpose of this paper, we do not attempt to build corrective models, but rather are interested in examining whether a simple classical measurement error model is accurate, or if in fact the differential measurement error might vary across time and treatment group in a longitudinal lifestyle intervention study.

1.4 Application

The existing measurement error literature has focused more on patient demographics such as body mass index, gender, or race and how these affect the amount of self-reported

measurement error (Neuhouser 2008, Mossavar-Rahmani 2017). These studies typically examine measurement error at one specific time point and/or a single observational cohort.

However, we can't assume these measurement error patterns remain constant in longitudinal lifestyle interventions. Self-reporting behaviors could change over time and/or by treatment assignment. Those in the treatment group may become more cognizant of nutrition intake through intervention exposure, leading to increased reporting accuracy. Participants may also modify their self-reported values (even if not necessarily their true intake) to appear compliant with intervention recommendations, which decreases their accuracy (Espeland et al. 2001).

Self-reported precision could also wane over time as participants experience a fatigue with repeated reporting (Buzzard 1996). This fatigue causes them to be more carefree and less rigorous, biasing results. Conversely, as people repeatedly monitor sodium intake over time, they may become more accurate with increased repetitions. Thus, the structure of the measurement error may change over time and by treatment group, an important consideration for measurement error correction. But to this point there has been little empirical investigation of these patterns.

2. DATA

Using two longitudinal lifestyle intervention trials, we examine sodium intake as a case study to learn about whether the measurement error structure differs across time and treatment group; the results can then be used to understand when it is important to consider differential measurement error by time or group.

To examine self-reported measurement error over time and across treatment groups, we used data from two longitudinal lifestyle intervention trials: Trials of Hypertension Prevention (TOHP) (Whelton et al. 1992) and PREMIER: Lifestyle Interventions for Blood Pressure Control (Appel et al. 2003).

These data sets are particularly useful for examining measurement error over time because they, unlike most nutrition trials, contain self-reported sodium intake and a sodium biomarker – 24-hour urine – for each participant and every time point. With this information, we compare the participants' self-reported values with their directly measured urinary sodium to characterize the measurement error, and assess whether the error varies across treatment group and time.

These internal validation datasets could be helpful to learn about potential measurement error in other settings, and to help researchers understand how much they may need to worry about differential measurement error across time and treatment.

2.1 Trials of Hypertension Prevention

TOHP was a US based, multicenter, randomized trial of 3 years duration with 2,182 participants testing the efficacy of a lifestyle intervention aimed at lowering diastolic blood pressure (DBP) in the high normal range (80 to 89 mmHg) (Satterfield et al. 1991). Participants were assigned to one of four treatment groups: Sodium reduction, weight reduction, stress management, or control. The sodium reduction group received counseling on how to reduce sodium consumption in everyday life. The weight reduction group received guidance on weight-loss techniques. The stress management group were provided coping mechanisms to handle stressful situations. The weight loss and stress management groups did not receive any counseling specifically on sodium intake. The control group did not receive any particular intervention or information; in this sense it was similar to a “usual care” condition.

Participants were considered eligible if they were healthy men and women, aged 30 through 54 years, who had high normal DBP and were not taking antihypertensive drugs for the prior 2 months (Satterfield et al. 1991). All participants were screened three times prior to enrollment to check eligibility requirements and then randomized to one of the four treatment groups. On the third screening, a 24 hour recall was conducted, and participants provided a 24-hour urine sample; this served as their “baseline” measurement. All participants were contacted again – at an unannounced point in time – 6 months and 18 months after enrollment to provide 24-hour food recall survey and 24 hour urine biomarker for sodium consumption at each respective time point.

2.2 PREMIER: Lifestyle Interventions for Blood Pressure Control

PREMIER was also a US based, multicenter randomized trial testing the effects of various lifestyle intervention on blood pressure outcomes in 810 individuals.

Participants were randomly assigned to one of three treatment groups: Established, Established Plus Dash, or Advice Only. The Established group received guidance on improving their dietary habits (including sodium consumption) and increase physical activity. Established Plus Dash received an intervention similar to Established but also received education on the DASH diet, a diet high in fruits, vegetables and low-fat dairy products. Finally, Advice Only received general healthy behavior advice, but no specific counseling on sodium intake or physical activity levels.

All eligible participants attended a randomization visit, where researchers randomized them to a group and then collected baseline measurements including two 24-hour diet recalls, and a 24-hour urine sample. Trial researchers contacted all participants unannounced at 6 and 18 months after enrollment, at which point individuals again provided two 24-hour diet recalls and 24-hour urine samples (Table 1).

We obtained the datasets for TOHP and PREMIER through an online request from the National Heart, Lung, and Blood Institute BioLINCC data repository after receiving IRB approval through Johns Hopkins Bloomberg School of Public Health.

For both datasets we consolidated the original treatment and control groups into new ones for our modelling purposes. In TOHP, only the sodium reduction group received counseling on sodium management. Hence, the sodium reduction cohort is the treatment group and the stress management, weight reduction, and original control cohorts were consolidated into a new control arm. For our data analyses using the PREMIER study, we included both behavioral intervention groups (Established, Established plus DASH) as the new treatment cohort, and used the advice only condition as the new control arm. We are interested in whether participants in the sodium

reduction interventions, more (or less) accurately reported their actual sodium intake compared to those in the advice only group.

Table 1: Study Characteristics

	TOHP	PREMIER
Enrollment Dates	1988-1990	1999-2001
Timing of Sodium Assessment	Baseline 6 months 18 months	Baseline 6 months 18 months
Assessment Method	24-hour recall 24-hour urine	Two 24-hour recalls 24-hour urine
Treatment Categories	Sodium Reduction* Weight Reduction` Stress Management` Control`	Established* Established Plus DASH* Advice Only`
N	2182	810
% Male	70 %	38 %
Mean Baseline BMI (sd)	27.6 (3.7)	33.1 (5.7)
Mean Baseline Age (sd)	43 (6.5)	50 (8.9)

Table 1: Provides overview of each study and baseline demographics.

*Categorized as “treatment” for our modeling purposes.

`Categorized as “control” for our modeling purposes.

3. METHODS

The same data cleaning procedures were used for both studies prior to analysis. First, the biomarker sodium values were converted to dietary sodium values by dividing urine sodium values by 0.86, as only 86% of sodium intake appears in urine (Holbrook et al. 1984, Willet 2013). The dietary sodium and self-reported sodium values were both log-transformed to make the respective distributions approximately normal. Finally, participants with extreme energy intakes (<500 kcal and >3,500 kcal for women, <800 kcal and >4000 kcal for men) (Holbrook et al. 1984, Willet 2013) were excluded for these analyses. The exclusion criteria eliminated 7(~1%) and 8(~1%) people from TOHP and PREMIER, respectively.

Mixed effects linear regression was used to estimate the relationship between log measured urine sodium and log self-reported sodium over time, separately by treatment group, and to account the correlation of measures within a participant over time. We centered log self-report [log self-report – mean log self-report at baseline] to easily compare each individual against the mean. We allow each individual to have a random intercept, and the (log centered) self-reported values to have a random slope, and used an unstructured covariance matrix. All missing data points were treated as “missing completely at random;” discussed further below. To estimate these models, we used the lme4 and lmerTest packages in R version 3.5.1

For each trial, we started with an initial model that includes both main effects for follow-up time, subjects self-reported intake and treatment, and two-way interactions between each pair, and interactions between all 3. This model also includes a random intercept and a random slope for subjects’ self-reported intake:

$$\begin{aligned}
U_{ij} = & \beta_0 + \beta_1 * self_{ij} + \beta_2 \mathbb{I}(time_j = 6) + \beta_3 \mathbb{I}(time_j = 18) + \beta_4 \mathbb{I}(time_j = 6) * TX_i \\
& + \beta_5 \mathbb{I}(time_j = 18) * TX_i + \beta_6 * self_{ij} \mathbb{I}(time_j = 6) + \beta_7 * self_{ij} \mathbb{I}(time_j = 18) \\
& + \beta_8 * self_{ij} \mathbb{I}(time_j = 6) * TX_i + \beta_9 * self_{ij} \mathbb{I}(time_j = 18) * TX_i \\
& + b_{0i} + b_{1i} * self_{ij} + e_{ij}
\end{aligned}$$

(1)

For each person i ($i = 1, \dots, N$), at time j ($j = \text{baseline}, 6 \text{ months}, 18 \text{ months}$), in our defined treatment group (TX ; $0 = \text{control}, 1 = \text{treatment}$) their urine measured – “true” – sodium intake is represented by U where $e \sim N(0, \sigma^2)$. Both U and $self_{ij}$ are log transformed values, and $self_{ij}$ is centered at their baseline measurements. $\mathbb{I}()$ is an indicator function which takes on either 0 or 1. b_{0i} is the random intercept and b_{1i} is the random slope for each person’s centered self-reported values respectively. We assume $b_{0i} \sim N(0, \tau_0^2)$ and $b_{1i} \sim N(0, \tau_1^2)$.

We excluded a main effect for treatment (TX) from the model because we assume treatment and control groups have similar sodium levels at baseline, at least in expectation (because of randomization).

Including the three-way (self-reported intake by time by treatment) interactions in this initial model allows the relationship between urine measured sodium intake and self-reported sodium changes over time in the control group and allows for urine measured sodium to differ between the treatment and control group over time. We include a time by treatment interaction to examine whether the control group and treatment group have different levels of urine measured sodium at 6 and 18 months.

A backwards variable selection was used to fit models to analyze these data. First, this initial saturated model with the three-way interaction (1) was fit. For all subsequent tests a significance level of 0.2 was used. We first tested the three-way interaction $self*time*treatment$. If at least one coefficient had a $p\text{-value} < 0.2$, we kept both interaction

terms in the model. If both coefficients had p-value > 0.2 , we dropped them from the model and fit the model described in (2), which omits the 3-way interaction.

The second potential model did not include the three-way interaction terms:

$$U_{ij} = \beta_0 + \beta_1 * self_{ij} + \beta_2 \mathbb{I}(time_j = 6) + \beta_3 \mathbb{I}(time_j = 18) + \beta_4 \mathbb{I}(time_j = 6) * TX_i \\ + \beta_5 \mathbb{I}(time_j = 18) * TX_i + \beta_6 * self_{ij} \mathbb{I}(time_j = 6) + \beta_7 * self_{ij} \mathbb{I}(time_j = 18) \\ + b_{0i} + b_{1i} * self_{ij} + e_{ij} \quad (2)$$

In model (2), we tested the significance of the self*time terms (β_6, β_7) , which examine whether the relationship between urine measured sodium and self-reported measures change over time, assuming any change is constant across the treatment and control groups. Once again, if both coefficients had p-values > 0.2 , we dropped them from the model and fitted our final model, Model (3):

$$U_{ij} = \beta_0 + \beta_1 * self_{ij} + \beta_2 \mathbb{I}(time_j = 6) + \beta_3 \mathbb{I}(time_j = 18) + \beta_4 \mathbb{I}(time_j = 6) * TX_i \\ + \beta_5 \mathbb{I}(time_j = 18) * TX_i + b_{0i} + b_{1i} * self_{ij} + e_{ij} \quad (3)$$

Model (3) allows urine measured sodium levels to change across time and treatment status. In this model we test the time*treatment interaction (β_4, β_5) . If both coefficients had p-values > 0.2 , we dropped them from the model.

We then standardized the regression coefficients using the mean and standard deviation of the pooled (control and treatment) group at baseline.

3.1 Covariate Modelling

While not the main purpose of this paper, we also investigated separately how BMI and gender influenced measurement error structure over time and by treatment assignment. Previous

research has shown these attributes factor into self-reported measurement error. We built similar models to those listed above, but included an additional BMI or gender interaction term with each previously included interaction, i.e., self*time*treatment*gender, self*time*gender. We then again applied the backwards variable selection process which dropped terms that are not statistically significant, at a 0.2 p-value threshold.

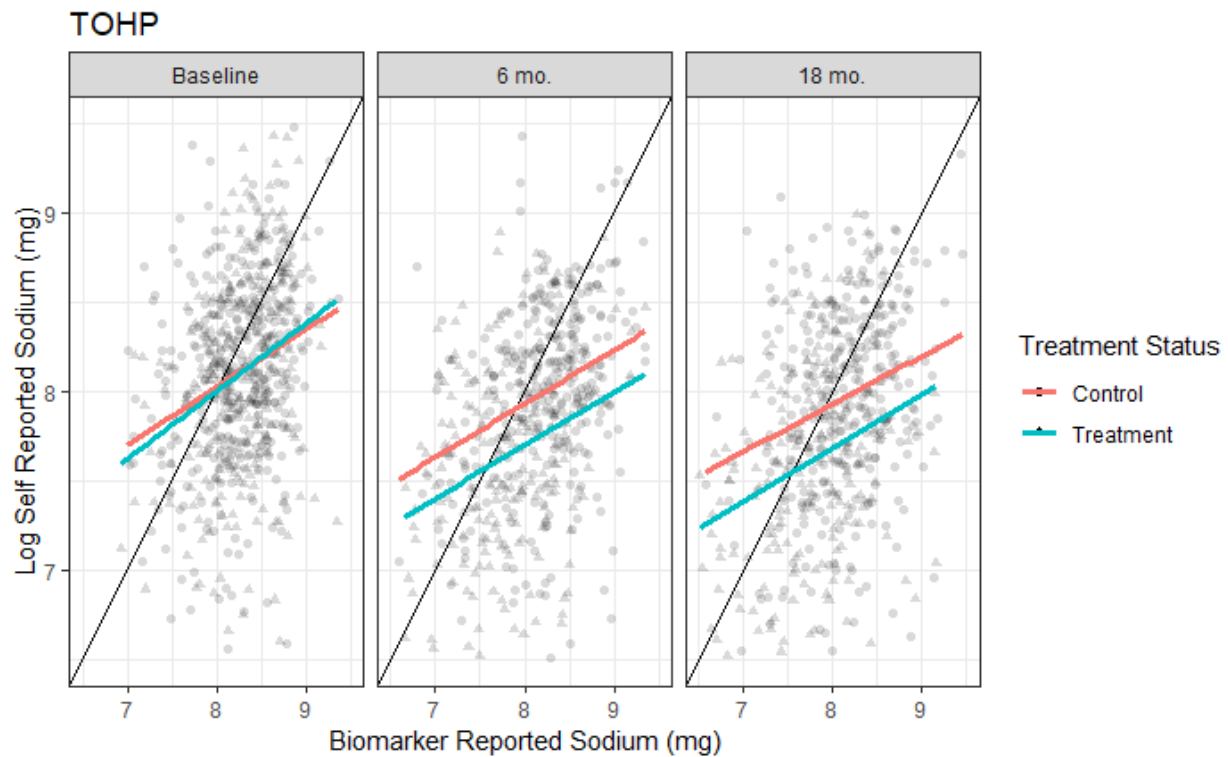
See SUPPLEMENTAIRES 6.1 for more detail

4. RESULTS

4.1 Descriptive Results

First, just examining the data descriptively by comparing the biomarker and self-reported values, both datasets include people who over and under report by time and treatment status (Figure 1). The 45-degree line in each graph represents “perfect” reporting, where measured urine biomarker equals self-reported sodium. Those who fall above the line over report, meaning their measured urine sodium levels were lower than self-reported levels. Conversely, those below the line under report, meaning their measured urine sodium levels were higher than their self-reported amounts. The wide scattering of points suggests a high degree of variability in reported sodium levels.

Figure 1: Measured Urine vs. Self-Reported intake in each study by time and treatment



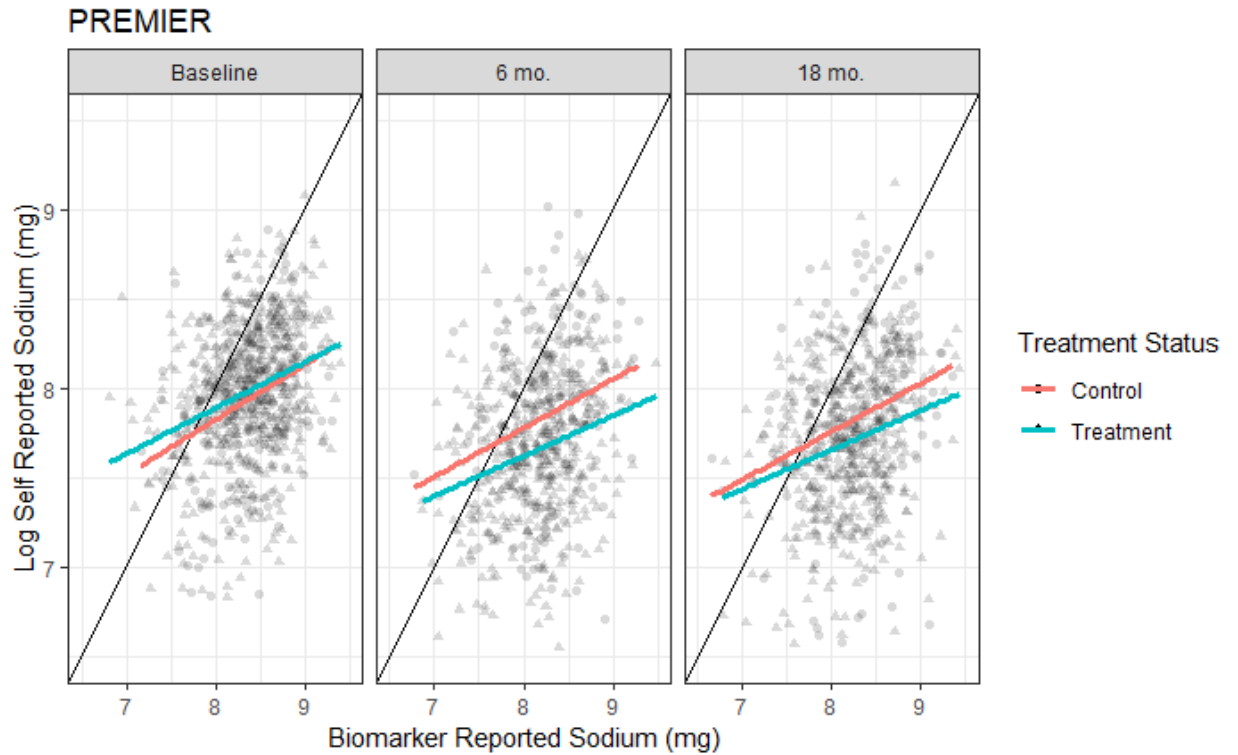


Figure 1: Provides data from both studies and a predicted self-reported line based on measured urine sodium by time and treatment status. 45-degree line represents where measured urine equals self-reported sodium.

Based on the data output from Figure 1, the linear predicted slopes are relatively equal for the treatment and control groups at baseline; a reasonable result since we assumed the groups were approximately equal at baseline. The correlation between self-reported sodium and measured urine sodium stays comparatively constant over time and by treatment assignment (Table2/Table3). The biggest discrepancy between intervention arms is the intercept location after baseline.

In the TOHP dataset, the control group and treatment group reported a mean measured urine sodium amount of 3827 mg of sodium at baseline. At 18 months, the control group averaged 364 mg less sodium compared to their baseline measurement, while the treatment group consumed 1506 mg less at 18 months compared to their average baseline intake (as measured by biomarker). Despite the difference in sodium intake between the two groups, the difference between self-reported sodium and measured urine biomarker at 18 months is very similar (a log difference of 0.19 and 0.2 respectively).

Table 2: Mean log values by time and treatment in the TOHP dataset

	Control			Treatment		
	Baseline	6 mo.	18 mo.	Baseline	6 mo.	18 mo.
Mean	8.25 (0.45)	8.24 (0.51)	8.15 (0.52)	8.25(0.40)	7.80 (0.54)	7.75 (0.57)
Urine (sd)						
Mean Self-Report (sd)	8.09 (0.54)	7.96 (0.54)	7.96 (0.59)	8.11 (0.54)	7.58 (0.58)	7.55(0.60)
Mean Difference	0.16	0.28	0.19	0.14	0.22	0.2
Correlation	0.28	0.28	0.27	0.27	0.31	0.34

Table 2: Provides the mean for measured urine and self-reported sodium, their difference, and correlation between the two measurements in TOHP dataset.

In the PREMIER dataset, the control group consumed an average 4359 mg of sodium, and the treatment group reported a mean of 4230 mg of sodium at baseline. At 18 months, the

control group averaged of 128 mg less sodium compared to their baseline measurement, while the treatment group consumed 661 mg less at 18 months compared to their average baseline intake (as measured by biomarker). Again, the difference between self-reported sodium and measured urine biomarker at 18 months is very similar (a log difference of 0.45 and 0.48 respectively). Overall, PREMIER reports higher averages of sodium intake than TOHP.

Table 3: Mean log values by time and treatment in PREMIER

	Control			Treatment		
	Baseline	6 mo.	18 mo.	Baseline	6 mo.	18 mo.
Mean	8.38 (0.40)	8.23 (0.44)	8.29 (0.46)	8.35 (0.45)	8.13 (0.54)	8.18 (0.47)
Urine (sd)						
Mean Self-Report (sd)	7.94 (0.39)	7.84 (0.45)	7.84 (0.44)	7.98 (0.39)	7.66 (0.41)	7.70 (0.42)
Mean Difference	0.44	0.39	0.45	0.37	0.47	0.48
Correlation	0.32	0.21	0.27	0.30	0.27	0.26

Table 2: Provides the mean for measured urine and self-reported sodium, their difference, and correlation between the two measurements in PREMIER dataset.

4.2 Regression Results

Neither the three-way interactions in model (1), nor the interactions between self-reported sodium levels and time in model (2) met the criteria for inclusion. As such, the final model only

includes the interaction between treatment and time (model 3). This model says average measured urine sodium changes over time (β_2, β_3), and at different rates in the treatment group vs. control group (β_4, β_5).

Table 4: Standardized Regression output from model (3)

	TOHP		PREMIER	
	Estimate (95%CI)	p-value	Estimate (95%CI)	p-value
Centered Self-Report β_1	0.29 (0.23,0.34)	<0.001*	0.21 (0.17, 0.26)	<0.001*
Time 6 Control β_2	0.03 (-0.10,0.16)	0.68	-0.24 (-0.37, -0.10)	<0.001*
Time 18 Control β_3	-0.19 (-0.32,-0.06)	0.005*	-0.08 (-0.21, 0.05)	0.23
Time 6 Trt. β_4	-0.81 (-1.0, -0.63)	<0.001*	-0.1 (-0.26, 0.06)	0.20*
Time 18 Trt. β_5	-0.65 (-0.84, -0.47)	<0.001*	-0.15 (-0.30, 0.0)	0.06*

Table 2: Provides standardized beta coefficient values controlling for time and treatment assignment, with their corresponding 95% confidence interval and p-values.

* Significant at 0.2 level

In TOHP, there was no significant interaction between average measured urine sodium and the difference between baseline and 6 months in the control group ($\beta_2 = 0.03$), for a given level of self-reported sodium. There was a significant decrease for the difference between baseline and 18 months however, in the control group ($\beta_3 = -0.19$). For the treatment interventions, there was on average a significant decrease in measured urine sodium between baseline and each follow up time, ($\beta_4 = -0.81, \beta_5 = -0.65$), for a given level of self-reported sodium.

In PREMIER, there was a significant decrease in average measured urine sodium at 6 months compared to baseline ($\beta_2 = -0.08$) in the control group, for a given level of self-report. There was no significant difference between 18 months and baseline in the control group ($\beta_3 = -0.08$). A significant decrease in average measured urine sodium occurred at 6 months ($\beta_4 = -0.1$) and 18 months ($\beta_5 = -0.15$) for the treatment category.

In both studies, the three-way interaction between time, treatment, and self-reported amounts were not significant, nor the two-way interaction between time and self-reported amount. This might indicate a lack of significant difference in systematic error between the treatment arms across all three time points.

These results appear consistent with Figure 1, as the average level of self-reported sodium doesn't seem to change dramatically based on treatment assignment or time point. However, the amount of reported measured urine sodium changes as a function of time, with a bigger decrease in average urine sodium in treatment group compared to the control group. TOHP reported less measured urine sodium than PREMIER consistent with Table 2/Table 3.

4.3 Covariate Results

For the gender and BMI covariate models, our outcomes are consistent with previous work and previous results of this paper. Both covariates affect the amount of predicted urine sodium and self-reported sodium levels.

We find that self-reported values do not change as a function of time, treatment, and gender/BMI in both studies. Consistent with model (3), expected measured sodium levels decrease over time, at different rates in the treatment vs. control group, even when controlling for

gender/BMI. Gender and BMI both affect self-reported values regardless of treatment or time status.

Overall, males over-report self-recorded sodium values, and have higher expected urine sodium levels. In PREMIER, males on average over reported sodium by 10% compared to females, and consumed 22% more sodium than females. In TOHP, males on average over reported by 8% compared to females, and consumed 22% more sodium. Both studies have imbalanced male to female ratio, and this could influence the results.

In BMI, we see expected urine sodium levels increase with higher BMI, and self-reported values decrease with higher BMI. On average as BMI increases, people consume more sodium, but self-report lower amounts, leading to under reporting. In PREMIER, each 1 unit increase in BMI lead to a 2% increase in predicted urine sodium, but a 1% decrease in self-reported value. In TOHP, a 1 unit increase in BMI cause a 3% increase in predict urine sodium. The self-reported*bmi coefficient is not significant at the 0.2 p-value threshold in TOHP.

See SUPPLEMENTARIES Section 7.2 For table output and graphs

5. DISCUSSION

Differential measurement error in nutrition studies may arise when the treatment group self-reports with increased or decreased accuracy of sodium intake (Sanjeevi 2019). Reporting bias can also occur over time in longitudinal studies, and may shift depending on treatment status.

Based on our regression modelling, participants' self-reported sodium levels do not significantly change as a function of time and/or treatment status. If this happened, either (β_8, β_9) , from model (1), the self-report*time*treatment status would be significant, or, (β_6, β_7) , from model (2), the self-report*time interaction.

However, from model (3) there is evidence of a relationship between sodium consumption and time/treatment status. Over time, lifestyle trial participants experience a decrease in sodium intake, with a greater reduction in urine measured sodium for the treatment group compared to the control group. If people perfectly reported their true intake, we would expect $\beta_1 = 1$ and $\beta_2, \beta_3, \beta_4, \beta_5 = 0$. In contrast, we find that $\beta_1 \neq 1$ and $\beta_2, \beta_3, \beta_4, \beta_5 < 0$, an indication measured urine sodium levels may change over time and by treatment status.

Discrepancies in the literature still exist about the relationship between treatment and self-reporting error. Even though this paper didn't find a relationship between treatment assignment and self-report bias, other studies have. In the Women's Health Eating and Living Study, a longitudinal randomized intervention trial with validation data (Natarajan et al. 2010), researchers found dietary intervention affected measurement error in self-reported outcomes using plasma carotenoid biomarkers. In the Women's Health Initiative Dietary Modification Trial, another dietary intervention trial (Neuhouser et al. 2008), participants in the control group under-reported protein intake at greater amounts compared to the treatment arm. There is thus

evidence that whether or not there is differential measurement error across time and treatment group may vary quite a bit across studies.

One possible solution to address measurement error would be more internal validation datasets with longitudinal intervention aspects. While this route is resource intensive, it may be worthwhile if researchers continue to study the relationship between disease outcomes and nutrient intake. Perhaps a cheaper or less invasive biomarker would make creating this dataset more feasible.

Another option would be more measurement error correction methods, which is why it is important to study how measurement error structures change over time and by treatment status. Siddique et al. (2018) use an assumption that the measurement error structure is time invariant, treatment invariant, and time and treatment invariant. Understanding how measurement errors structures change as a function of treatment assignment and time point might improve the robustness of new methods, especially when trying to generalize the results of a study to the population. Documenting time and treatment effects with regression calibration in internal validation datasets might improve the accuracy of self-reported measures in longitudinal intervention trials without available biomarker data.

5.1 Limitations

One large limitation of this study is the amount of missing data, as high as 28% by 18 months in TOHP (Table 3). We assumed missing at random, but this is an extreme assumption given the nature of the data – a longitudinal intervention trial. The assumption is likely not fully accurate as people can voluntarily remove themselves from the trial for various reasons. Participants who drop out likely behave differently than those who stay in the trial, creating bias in final results.

Table 5: Percent Missing by Time/Treatment in each study

	TOHP		PREMIER	
	Control	Treatment	Control	Treatment
Baseline	1	9.5	1	9
6 months	23	29.3	23	19.6
18 months	18.6	27.8	15	18.7

Table 3: Percent missing by treatment category at each timepoint in TOHP and PREMIER.

Additionally, both biomarker and self-reported methods experience error which can affect the measured intake levels. It might be difficult to quantify the amount of systematic error within the dataset and thus might be hard to establish how time and treatment status affect the true relationship between biomarker and self-reported data.

Finally, it's important to know that a single intake measurement, whether its biomarker or self-reported, may not be representative of a person's diet (Kahn 1995). Collecting only three measurement over the course of 18 months might not truly capture dietetic patterns, especially if the fluctuate in individuals or by seasons.

Documenting the measurement error structure by time and treatment in longitudinal studies is important if researchers want to properly measure the relationship between nutrient intake and disease outcome. If self-reporting habits changed based on time or treatment status, then lifestyle intervention trials that fail to account for this, may draw erroneous conclusions of their results. Additionally, if researchers attempted to improve their measurement error models but don't include important covariates, like BMI, gender, time and/or treatment, this may cause correction issues in the statistical model, once again leading to a possibly incorrect relationship of exposure and outcome.

6. SUPPLEMENTARIES

6.1 Covariate Models

Listed below are models (1), (2), (3) for sex. Now we include an additional indicator - $\mathbb{I}()$ - with subscript k indicating sex (k ; 0 = Female, 1 = male).

$$\begin{aligned}
 U_{ijk} = & \beta_0 + \beta_1 * \text{self}_{ijk} + \beta_2 \mathbb{I}(\text{time}_j = 6) + \beta_3 \mathbb{I}(\text{time}_j = 18) + \beta_4 \mathbb{I}(\text{time}_j = 6) * \text{TX}_i \\
 & + \beta_5 \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_6 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) + \beta_7 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) \\
 & + \beta_8 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i + \beta_9 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_{10} \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{11} * \text{self}_{ijk} \mathbb{I}(\text{male}_k = 1) + \beta_{12} \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{13} \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) + \beta_{14} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{15} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{16} * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{17} * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{18} * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{19} * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) + b_{0i} + b_{1i} * \text{self}_{ij} + e_{ij}
 \end{aligned}
 \tag{1}$$

$$\begin{aligned}
 U_{ijk} = & \beta_0 + \beta_1 * \text{self}_{ijk} + \beta_2 \mathbb{I}(\text{time}_j = 6) + \beta_3 \mathbb{I}(\text{time}_j = 18) + \beta_4 \mathbb{I}(\text{time}_j = 6) * \text{TX}_i \\
 & + \beta_5 \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_6 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) + \beta_7 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) \\
 & + \beta_8 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i + \beta_9 * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_{10} \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{11} * \text{self}_{ijk} \mathbb{I}(\text{male}_k = 1) + \beta_{12} \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{13} \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) + \beta_{14} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{15} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{16} * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) \\
 & + \beta_{17} * \text{self}_{ijk} \mathbb{I}(\text{time}_j = 18) * \mathbb{I}(\text{male}_k = 1) \\
 & + b_{0i} + b_{1i} * \text{self}_{ij} + e_{ij}
 \end{aligned}
 \tag{2}$$

$$\begin{aligned}
U_{ijk} = & \beta_0 + \beta_1 * \text{self}_{ijk} + \beta_2 \mathbb{I}(\text{time}_j = 6) + \beta_3 \mathbb{I}(\text{time}_j = 18) + \beta_4 \mathbb{I}(\text{time}_j = 6) * \text{TX}_i \\
& + \beta_5 \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_6 \mathbb{I}(\text{male}_k = 1) \\
& + \beta_7 * \text{self}_{ijk} \mathbb{I}(\text{male}_k = 1) + \beta_8 \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) \\
& + \beta_9 \mathbb{I}(\text{time}_j = 6) * \mathbb{I}(\text{male}_k = 1) + \beta_{10} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
& + \beta_{11} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \mathbb{I}(\text{male}_k = 1) \\
& + b_{0i} + b_{1i} * \text{self}_{ij} + e_{ij}
\end{aligned}
\tag{3}$$

For BMI, we include a subscript l to indicate a participant's BMI (continuous) value. BMI was centered with respected to all (both treatment and control) participant's baseline BMI level.

$$\begin{aligned}
U_{ijl} = & \beta_0 + \beta_1 * \text{self}_{ijl} + \beta_2 \mathbb{I}(\text{time}_j = 6) + \beta_3 \mathbb{I}(\text{time}_j = 18) + \beta_4 \mathbb{I}(\text{time}_j = 6) * \text{TX}_i \\
& + \beta_5 \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_6 * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 6) + \beta_7 * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 18) \\
& + \beta_8 * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i + \beta_9 * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_{10} * \text{BMI}_l \\
& + \beta_{11} * \text{self}_{ijl} \mathbb{I} * \text{BMI}_l + \beta_{12} \mathbb{I}(\text{time}_j = 6) * \text{BMI}_l \\
& + \beta_{13} \mathbb{I}(\text{time}_j = 18) * \text{BMI}_l + \beta_{14} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i * \text{BMI}_l \\
& + \beta_{15} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \text{BMI}_l \\
& + \beta_{16} * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 6) * \text{BMI}_l \\
& + \beta_{17} * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 18) * \text{BMI}_l \\
& + \beta_{18} * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i * \text{BMI}_l \\
& + \beta_{19} * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \text{BMI}_l + b_{0i} + b_{1i} * \text{self}_{ij} + e_{ij}
\end{aligned}
\tag{1}$$

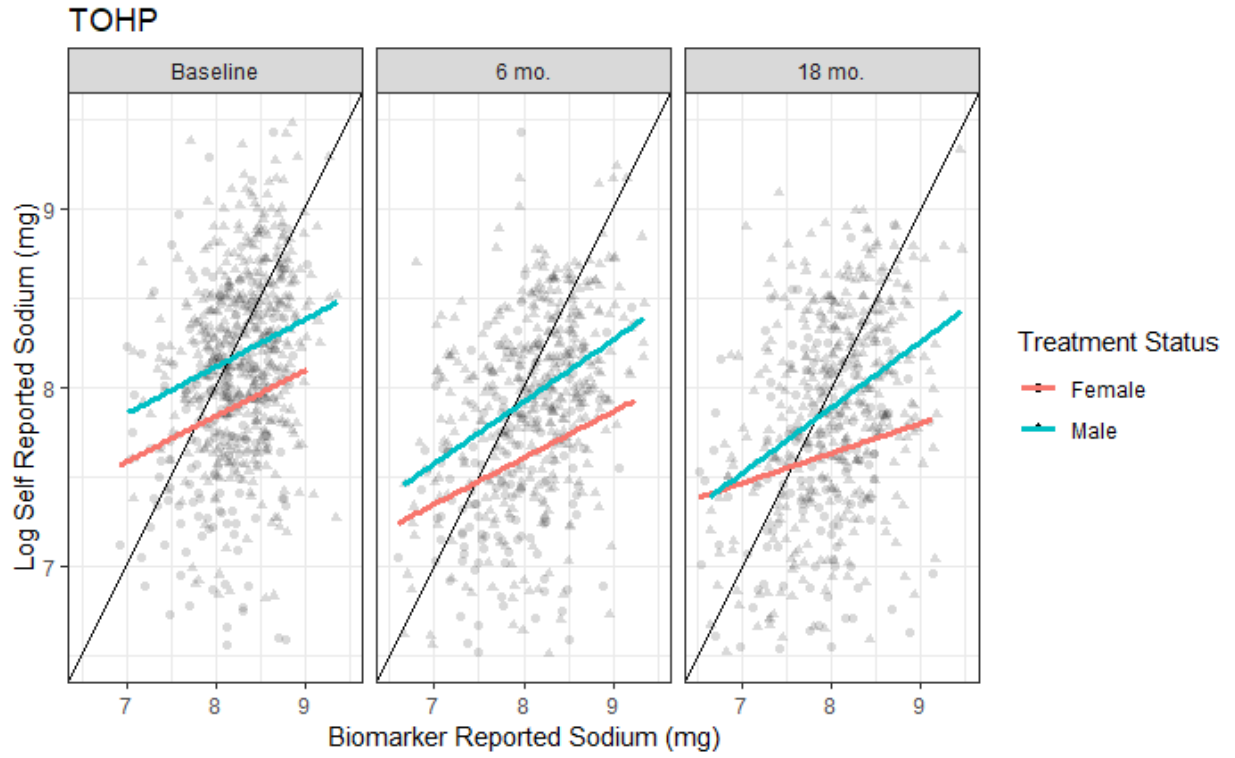
$$\begin{aligned}
U_{ijl} = & \beta_0 + \beta_1 * \text{self}_{ijl} + \beta_2 \mathbb{I}(\text{time}_j = 6) + \beta_3 \mathbb{I}(\text{time}_j = 18) + \beta_4 \mathbb{I}(\text{time}_j = 6) * \text{TX}_i \\
& + \beta_5 \mathbb{I}(\text{time}_j = 18) * \text{TX}_i + \beta_6 * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 6) + \beta_7 * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 18) \\
& + \beta_8 * \text{BMI}_l + \beta_9 * \text{self}_{ijl} * \text{BMI}_l + \beta_{10} \mathbb{I}(\text{time}_j = 6) * \text{BMI}_l \\
& + \beta_{11} \mathbb{I}(\text{time}_j = 18) * \text{BMI}_l + \beta_{12} \mathbb{I}(\text{time}_j = 6) * \text{TX}_i * \text{BMI}_l \\
& + \beta_{13} \mathbb{I}(\text{time}_j = 18) * \text{TX}_i * \text{BMI}_l \\
& + \beta_{14} * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 6) * \text{BMI}_l \\
& + \beta_{15} * \text{self}_{ijl} \mathbb{I}(\text{time}_j = 18) * \text{BMI}_l \\
& + b_{0i} + b_{1i} * \text{self}_{ij} + e_{ij}
\end{aligned}
\tag{2}$$

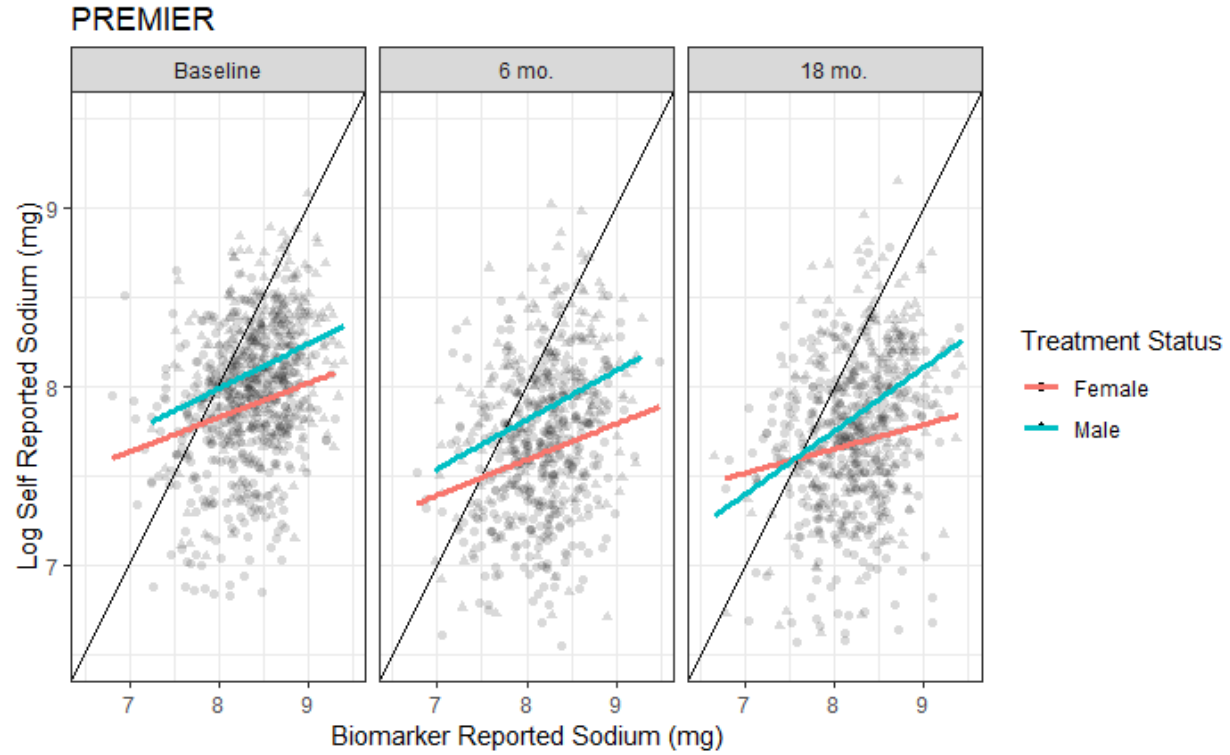
$$\begin{aligned}
U_{ijl} = & \beta_0 + \beta_1 * self_{ijl} + \beta_2 \mathbb{I}(time_j = 6) + \beta_3 \mathbb{I}(time_j = 18) + \beta_4 \mathbb{I}(time_j = 6) * TX_i \\
& + \beta_5 \mathbb{I}(time_j = 18) * TX_i + \beta_6 * BMI_1 + \beta_7 * self_{ijl} * BMI_1 \\
& + b_{0i} + b_{1i} * self_{ij} + e_{ij}
\end{aligned}$$

(3)

6.2 Covariate Output

Supplementary Figure 1: Measured urine vs. self-report by sex in both datasets





Supp. Figure 1: Descriptive data comparing males and females with their measured urine vs. self-reported values in both datasets.

Supplementary Table 1: Standardized Regression output from BMI model (3)

	TOHP		PREMIER	
	Estimate (95%CI)	p-value	Estimate (95%CI)	p-value
Centered Self-Report β_1	0.28 (0.23,0.34)	<0.001*	0.20 (0.16, 0.25)	<0.001*
Time 6 Control β_2	0.03 (-0.10,0.16)	0.63	-0.24 (-0.37, -0.10)	<0.001*
Time 18 Control β_3	-0.17 (-0.30,-0.04)	0.01*	-0.07 (-0.20, 0.06)	0.26
Time 6 Trt. β_4	-0.82 (-1.0, -0.64)	<0.001*	-0.11 (-0.27, -0.05)	0.16*
Time 18 Trt. β_5	-0.68 (-0.86, -0.49)	<0.001*	-0.16 (-0.31, -0.01)	0.04*
Centered BMI β_6	0.28 (0.23, 0.34)	<0.001*	0.25 (0.19, 0.30)	<0.001*

Centered Self * BMI β_7	0.02 (-0.03, 0.07)	0.5	-0.04 (-0.08,0.0)	0.07*
-------------------------------------	--------------------	-----	-------------------	-------

Table 1: Provides standardized beta coefficient values controlling for time and treatment assignment, and sex, with their corresponding 95% confidence interval and p-values.

* Significant at 0.2 level

Supplementary Table 2: Standardized Regression output from sex model (3)

	TOHP		PREMIER	
	Estimate (95%CI)	p-value	Estimate (95%CI)	p-value
Centered Self-Report β_1	0.18 (0.07,0.28)	<0.001*	0.16 (0.10, 0.21)	<0.001*
Time 6 Control β_2	-0.03 (-0.28,0.22)	0.80	-0.18 (-0.36, -0.01)	0.04*
Time 18 Control β_3	-0.16 (-0.41,0.09)	0.22	-0.11 (-0.28, 0.05)	0.18*
Time 6 Trt. β_4	-0.74 (-1.10, -0.38)	<0.001*	-0.01 (-0.21, 0.09)	0.93
Time 18 Trt. β_5	-0.75 (-1.10, -0.39)	<0.001*	-0.01 (-0.20, 0.18)	0.92
Male β_6	0.48 (0.29, 0.66)	<0.001*	0.46 (0.31, 0.60)	<0.001*
Centered Self * Male β_7	0.11 (-0.02, 0.23)	0.09*	0.07 (-0.02,0.16)	0.15*
Male *Time 6 Control β_8	0.06 (-0.23,0.35)	0.69	-0.19 (-0.47, 0.09)	0.19*
Male *Time 18 Control β_9	-0.07 (-0.37,0.22)	0.63	0.06 (-0.20, 0.33)	0.63

Male * Time 6 Trt. β_{10}	-0.13 (-0.55, 0.29)	0.54	-0.26 (-0.59, -0.07)	0.12*
Male * Time 18 Trt. β_{11}	0.10 (-0.32, 0.52)	0.66	-0.38 (-0.69, -0.07)	0.01*

Table 2: Provides standardized beta coefficient values controlling for time and treatment assignment, and sex, with their corresponding 95% confidence interval and p-values.

* Significant at 0.2 level

7. REFERENCES

- Alexander, L. Lopes, B. Ricchetti-Masterson, K. Yeatts, K. (2015). Sources of Systematic Error or Bias: Information Bias. *University of North Carolina: Chapel Hill*. Department of Epidemiology. 2nd Edition.
- Appel, L. J., et al. (2003). Effects of comprehensive lifestyle modification on blood pressure control: Main results of the PREMIER clinical trial. *JAMA* 289, 2083-2093.
- Buzzard, et al. (1996). Monitoring dietary change in a low-fat diet intervention study: Advantages of using 24-hour dietary recalls vs food records. *Journal of the American Dietetic Association* 96, 574 - 579.
- Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006). Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. Chapman & Hall/CRC, New York.
- Espeland, M. A. et al. (2001). Lifestyle interventions influence relative errors in self-reported diet intake of sodium and potassium. *Annals of Epidemiology* 11, 85 - 93.
- Gibson, R. Charrondiere, U. Bell, W. (2017). Measurement Errors in Dietary Assessment Using Self-Reported 24-Hour Recalls in Low-Income Countries and Strategies for Their Prevention. *Advances in Nutrition*. 8(6): 980–991.
- Holbrook, J.T., Patterson, K.Y., Bodner, J.E., Douglas, L.W., Veillon, C., Kelsay, J.L., Mertz, W., Smith, J.C. (1984). Sodium and potassium intake and balance in adults consuming self-selected diets. *The American Journal of Clinical Nutrition* 40:786-793.
- Kahn et al (1995). Validity of 24-Hour Dietary recall interviews conducted among volunteers in an adult working community. *Annals of Epidemiology*. 5,6,484-489.
- Keogh R, Carroll R, Tooze J, Kirkpatrick S, Freedman L (2016). Statistical issues related to dietary intake as the response variable in intervention trials. *Statistics in Medicine*. 35, 25, 4493–4508.
- Kirkpatrick S, Collins C, Keogh R, Krebs-Smith S, Neuhaus M, Wallace A, (2018) Assessing dietary outcomes in intervention studies: Pitfalls, strategies, and research needs. *Nutrients*. 10, 1001
- Mossavar-Rahmani et al. (2017). Applying recovery biomarkers to calibrate self-report measures of sodium and potassium in the Hispanic Community Health Study/Study of Latinos. *Journal of Human Hypertension*. 31, 462-473

- Natarajan, L. et al. (2010). Measurement error of dietary self-report in intervention trials. *American Journal of Epidemiology* 172, 819 - 827.
- National Institutes of Health. (2014) Dietary Assessment Primer, Regression Calibration. National Institutes of Health. <https://dietassessmentprimer.cancer.gov/>. Accessed on June 22, 2019.
- Neuhouser M, et al. (2008). Use of Recovery Biomarkers to Calibrate Nutrient Consumption Self-Reports in the Women's health Initiative. *American Journal of Epidemiology*. Vol 167, 10, 15. 1247-1259.
- Neuhouser M. (2017). Overview of the State of Biomarkers of Nutrient Intake. *Fred Hutchinson Cancer Research Center*. Seattle, WA. Presented on January 9, 2017. http://nationalacademies.org/hmd/~media/Files/Activity%20Files/Nutrition/Framework%20DRI/1_Nutritional%20Biomarkers%2001-09-2017%20V2.pdf. Retrieved on June 23, 2019
- Prentice, R. Sugar, E. Wang, C. Neuhouser, M. Patterson, R. (2002). Researcher strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Public Health Nutrition*. 5, 6a: 977-984. DOI: <https://doi.org/10.1079/PHN2002382>
- Sanjeevi N, Lipsky L, Nansel T. (2019). Differential reporting of fruit and vegetable intake among youth in a randomized controlled trial of a behavioral nutrition intervention. *International Journal of Behavioral Nutrition and Physical Activity*. 16, 15.
- Shim J, Oh K, Kim H (2014), Dietary assessment methods in epidemiologic studies. *Epidemiology and Health*. 36.
- Thompson FE, Kirkpatrick SI, Subar AF, Reedy J, Schap TRE, Wilson M, Krebs-Smith SM. (2015). The National Cancer Institute's Dietary Assessment Primer: a resource for diet research. *Advances in Nutrition* Diet 115:1986–95. DOI: 10.3945/an.117.016980
- Whelton P, Appel L, Charleston H. (1992) The effects of nonpharmacologic interventions on blood pressure of persons with high normal levels. Results of the Trials of Hypertension Prevention, Phase I. *Journal of the American Medical Association*. 267,9: 1213-1220. DOI :10.1001/jama.1992.03480090061028
- Willet, W. (2013). *Nutritional Epidemiology*, Third Edition. Oxford University Press, New York.

CURRICULUM VITAE

EDUCATION

Master of Science (ScM) in Biostatistics Expected August 2019
Johns Hopkins Bloomberg School of Public Health Baltimore, MD

Bachelor of Arts (BA) in Public Health with Honors December 2016
University of Colorado Denver Denver, CO

ANALYTIC EXPERIENCE

Johns Hopkins Bloomberg School of Public Health June 2018 - Present
Research Assistant | Measurement Error and Casual Inference Group Baltimore, MD

- Reviewing 87 published papers to improve current mediation analysis in mental health research
- Building 12 regression models in R to assess measurement error in self-reported sodium intake data

Center of Aging and Health September 2017 - September 2018
Research Fellow | Department of Medicine
Baltimore, MD

- Predicted four cognitive outcomes and four physical outcomes in 600+ adults using 100,000+ healthcare dataset
- Utilized random forest, k-means clustering, and logistic regression techniques in high-dimensional medical data to analyze physical performance in R
- Provided three intervention recommendations to improve physical activity in older adults with vision impairment

Harm Reduction Action Center
October 2016 - Jan 2017
Volunteer Data Analyst | Syringe Access Program
Denver, CO

- Educated groups of four to eight intravenous drug users on 10+ harm reduction behaviors to mitigate damage
- Analyzed data in SAS of 300+ patients collected first hand for 22-page honors capstone project
- Summarized six core findings from CDC dataset to present to Colorado legislature encouraging syringe access programs

WORK EXPERIENCE

Teaching Assistant June 2018 - Present
Johns Hopkins Bloomberg School of Public Health Baltimore, MD

- Courses: *Advanced Data Analysis Workshop, Statistical Methods in Public Health, Statistical Reasoning*

SELECTED PRESENTATIONS

- How providing clean needles protects the community December 2016
Colorado School of Public Health. Denver, CO

PROFESSIONAL DEVELOPMENT

Technologies: R, STATA, SAS, SQL, Python, MS Office, GitHub

Memberships: American Statistical Association (ASA). Student Outreach Resource Center (SOURCE). JHSPH Biostatistics Journal and Computing Club. Johns Hopkins Consulting Club.

Certifications: Data Science Specialization in R (DataCamp). SQL Certification (Coursera)